What is claimed is:

1. A method of modeling a directed evolution protocol comprising:
applying equilibrium thermodynamics to a plurality of sequences to determine statistics of
    hybridization; and
parameterizing an assembly algorithm using the statistics of hybridization.

2. The method of claim 1 further comprising applying the assembly algorithm to reassemble a plurality of sequences.

3. The method of claim 2 further comprising determining crossover allocation in the plurality of reassembled sequences.

4. The method of claim 3 wherein the step of determining crossover allocation includes estimating a fraction of the plurality of reassembled sequences containing a number of crossovers.

5. The method of claim 3 wherein the step of determining crossover allocation includes estimating a probability that a given nucleotide position in one of the plurality of reassembled sequences is a site of a crossover event.

6. The method of claim 1 wherein the directed evolution protocol is DNA shuffling.

7. The method of claim 1 wherein the directed evolution protocol is SCRATCHY.

8. The method of claim 1 further comprising identifying a minimum number of required silent mutations to meet a DNA recombination objective.

9. The method of claim 1 wherein the step of applying equilibrium thermodynamics to determine statistics of hybridization includes:
modeling annealing events during reassembly as a network of reactions;
determining a predicted fraction of fragments that will anneal at a given temperature;

36

determining a predicted distribution of annealing for overlap lengths; and

determining a portion of annealing events predicted to involve mismatches.

10. The method of claim 1 wherein the assembly algorithm excludes silent crossovers.

11. An isolated nucleic acid molecule comprising:

a nucleotide sequence having an amino acid sequence;

the nucleotide sequence isolated at least in part through a directed evolution experiment;
   and

the directed evolution experiment selected at least in part by applying equilibrium
   thermodynamics to a plurality of sequences to determine statistics of hybridization
   and parameterizing an assembly algorithm using the statistics of hybridization.

12. A vector comprising the nucleic acid molecule of claim 11.

13. A host cell containing the vector of claim 12.

14. A protein encoded by the nucleic acid sequence of claim 11.

15. A system for modeling a directed evolution protocol comprising:

a plurality of sequences; and

an article of software for determining statistics of hybridization of the plurality of
   sequences to parameterize an assembly algorithm by applying equilibrium
   thermodynamics to the plurality of sequences.